Are Options on Index Futures Profitable for Risk Averse Investors? Empirical Evidence

George M. Constantinides
University of Chicago and NBER

Michal Czerwonko
Concordia University

Jens Carsten Jackwerth
University of Konstanz

Stylianos Perrakis
Concordia University

Abstract

American call and put options on the S&P 500 index futures that violate the stochastic dominance bounds of Constantinides and Perrakis (2007) over 1983-2006 are identified as potentially profitable investment opportunities. Call bid prices more frequently violate their upper bound than put bid prices do, while evidence of underpriced calls and puts over this period is scant. In out-of-sample tests, the inclusion of short positions in such overpriced calls, puts, and, particularly, straddles in the market portfolio is shown to increase the expected utility of any risk averse investor and also increase the Sharpe ratio, net of transaction costs and bid-ask spreads. The results are strongly supportive of mispricing. (JEL G11, G13, G14)

Current draft: March 16, 2008

Keywords: option mispricing; futures options; derivatives pricing; stochastic dominance; transaction costs; market efficiency

We thank Russell Davidson for insightful comments and constructive criticism. We remain responsible for errors and omissions. Constantinides acknowledges financial support from the Center for Research in Security Prices of the University of Chicago; Czerwonko and Perrakis acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. E-mail addresses: gmc@ChicagoGSB.edu, m_czerwo@jmsb.concordia.ca, Jens.Jackwerth@uni-konstanz.de, SPerrakis@jmsb.concordia.ca

A large body of finance literature addresses the mispricing of options. Rubinstein (1994), Jackwerth and Rubinstein (1996), and Jackwerth (2000), among others, observed a steep index smile in the implied volatility of S&P 500 index options that suggests that out-of-the-money (OTM) puts are too expensive. Indeed, a common hedge-fund policy is to sell OTM puts. Coval and Shumway (2001) found that buying zero-beta at-the-money (ATM) straddles loses money. Constantinides, Jackwerth, and Perrakis (2008) provided empirical evidence that both OTM puts and calls on the S&P 500 index are mispriced by showing that they violate stochastic dominance bounds put forth by Constantinides and Perrakis (2002).

In this paper, we provide out-of-sample tests of option mispricing, net of transaction costs and bid-ask spreads. Specifically, we identify American call and put options on the S&P 500 index futures that violate the stochastic dominance bounds of Constantinides and Perrakis (2007) as potentially profitable investment opportunities. In out-of-sample tests over 1983-2006, we show that trading policies that exploit these violations provide higher Sharpe ratios than policies without option trading. We also show that the expected utility of any risk averse investor, net of transaction costs and bid-ask spreads, increases when exploiting such option trading. Below, we highlight novel features of our approach.

First, we use the Chicago Mercantile Exchange (CME) data base on S&P 500 futures options, 1983-2006, which is clean and spans a long period. Much of the earlier empirical work on the mispricing of index options is based on data on the S&P 500 index options that comes from two principal sources: the Berkeley Options Database (1986-1995) that provides relatively clean transaction prices, but misses important events over the past 12 years, such as the 1998 liquidity crisis, the dot-com bubble, and its 2000 burst; and the OptionMetrics (1996-2006) data base which, however, is of uneven quality and contains only end-of-day quotes.

Second, we identify mispriced options with a screening mechanism that uses *minimal* assumptions about market equilibrium. This mechanism is based on the stochastic dominance bounds of Constantinides and Perrakis (2007). These bounds identify reservation purchase and reservation write prices such that *any* risk averse investor may increase her expected utility by including the option that violates these bounds in her portfolio. The bounds are valid for any distribution of the underlying asset and accommodate jumps. They also recognize the possibility of early exercise of American options.

The only necessary assumption about the market for the validity of these bounds is that there exists a class of traders holding portfolios containing only the S&P 500 index and the riskless asset.[1] Ample evidence exists that this assumption holds for US markets. Numerous surveys have shown that a large number of US investors follow indexing policies in their investments. Bogle (2005) reports that in 2004 index funds accounted for about one third of equity fund cash inflows since 2000 and represented about one seventh of equity fund assets. The S&P 500 index is not only the most widely quoted market index, but has also been available to investors through exchange traded funds for several years. We find that any such investor would improve her utility by including in her portfolio an option identified as mispriced by the stochastic dominance bounds.

As a third novel feature, we assess the profitability of our trading policy by employing the powerful statistical tests of stochastic dominance by Davidson and Duclos (2000 and 2006) which can deal with option returns even in a setting where we do not make assumptions about the preferences of the investors. These tests compare the profitability of the optimal trading policies of a generic S&P 500 index investor with and without the option in a setting that recognizes the possibility of early exercise of the futures option. These profitability comparisons are valid from the perspective of any risk averse investor. By contrast, the ubiquitous Sharpe ratio measure of portfolio performance is valid only from the perspective of a mean-variance investor and suffers from well known problems when used to assess non-normal returns such as those encountered in portfolios that include options.

Finally, both the bounds employed in detecting mispriced options and the portfolio returns explicitly take into consideration bid-ask spreads and trading costs. Once a trading opportunity is detected, we execute the trade by buying at the next ask price or selling at the next bid price.

We use historical data on the underlying S&P 500 index returns in order to estimate the bounds. We use several empirical estimates of the underlying return distribution, all of them observable at the time the trading policy is implemented. For each one of these

[1] The mean-variance portfolio theory that gives rise to the Sharpe ratio measure of portfolio performance is based on the stronger assumption that every investor holds the market portfolio and the risk free asset.

estimates we evaluate the corresponding bounds over the period 1983-2006, and then identify the observed S&P 500 futures options prices that violate them. For each violation, we identify the optimal trading policy of a generic investor with and without the mispriced option, using the observed path of the underlying asset till option expiration and recognizing realistic trading conditions such as possible early exercise and transaction costs. We identify the profitability of the pair of policies for each observed violation, and then conduct stochastic dominance comparison tests over the entire sample of violations.

We find a substantial number of violations of the upper bounds, but relatively few violations of the lower bounds. Since the frequency of violations of the lower bounds is too low for statistical inference, we focus on violations of the upper bounds. The results are strongly supportive of mispricing.

The paper is organized as follows. In Section 1, we present the restrictions on futures option prices imposed by stochastic dominance and discuss the underlying assumptions. In Section 2, we describe the data and the empirical design. In Section 3, we present the empirical results and discuss their robustness. We conclude in Section 4.

1 Restrictions on Futures Option Prices Imposed by Stochastic Dominance

We assume that market agents are heterogeneous and investigate the restrictions on option prices imposed by one particular class of agents that we simply refer to as "traders". We allow for other agents to participate in the market but this allowance does not invalidate the restrictions on option prices imposed on traders.

We consider a market with several types of financial assets. First, we assume that traders invest only in two of them, a bond and a stock with natural interpretation as a market index.[2] Subsequently, we assume that traders can invest in a third asset as well, an American call or put option on the index futures. The bond is risk free and has total return

[2] Essentially, we model buy-and-hold investors who trade infrequently and incur low transaction costs. At least for large investors who earn a fair return on their margin, transaction costs are even lower in the index futures market than the stock market. In practice, however, buy-and-hold investors invest in the stock and bond markets because of the inconvenience and cost of the frequent rolling over of short-term futures contracts and the illiquidity of long-term futures and forward contracts.

R. The stock has *ex dividend* stock price S_t at time t and pays cash dividend γS_t, where the dividend yield γ is deterministic. The total return on the stock, $(1+\gamma)\left(S_{t+1}/S_t\right)$, is assumed *i.i.d.* with mean R_S. The call or put option on the index futures has strike K and expiration date T. The underlying futures contract is cash-settled and has maturity T^F, $T^F \ge T$. We assume that the futures price F_t is linked to the stock price by the approximate cost-of-carry relation $F_t = \left(1+\gamma\right)^{-\left(T^F - t\right)} R^{T^F - t} S_t + \varepsilon_t, \quad t \le T^F, \quad \left|\varepsilon_t\right| \le \overline{\varepsilon}$, where the basis risk variables $\left\{\varepsilon_t\right\}$ are distributed independently of each other and of the stock price series $\left\{S_t\right\}$.

Transfers to and from the cash account (bond trades) do not incur transaction costs. Stock trades decrease the bond account by transaction costs equal to the absolute value of the dollar transaction, times the proportional transaction costs rate, $k, 0 \le k < 1$. Option trades incur transaction costs, exchange fees, and price impact which are incorporated in what we refer to as their bid and ask prices.

We assume that traders maximize generally heterogeneous, state-independent, increasing, and concave utility functions. We further assume that each trader's wealth at the end of each period is monotone increasing in the stock return over the period. For example, a trader who holds 100 shares of stock and a net short position in 200 call options violates the monotonicity condition, while a trader who holds 200 shares of stock and a net short position in 200 call options satisfies the condition. Essentially, we assume that the traders have a sufficiently large investment in the stock, relative to their net short position in call options (or, net long positions in put options), such that the monotonicity condition is satisfied.

We do not make the restrictive assumption that all market agents belong to the class of utility-maximizing traders. Thus, our results are robust and unaffected by the presence in the market of agents with beliefs, endowments, preferences, trading restrictions, and transaction costs schedules that differ from those of the utility-maximizing traders modeled in this paper.

A trader enters the market at time zero with x_0 dollars in bonds and y_0 dollars in *ex dividend* shares of stock. We consider two scenarios. In the first scenario, the trader may trade the bond and stock but not the options. The trader makes sequential investment decisions at discrete trading dates t ($t = 0, 1\ldots, T'$), where $T', T' \geq T^F \geq T$, is the finite terminal date. The trader's objective is to maximize expected utility, $E[u_{T'}(W_{T'})]$, where $W_{T'}$ is the trader's net worth at date T'.[3] Utility is assumed to be concave and increasing and defined for both positive and negative terminal worth, but is otherwise left unspecified. We refer to this trader as the *index* (and bond) trader, IT, and denote her maximized expected utility by $V_0^{IT}(x_0, y_0)$.

In the second scenario, as in the first scenario, the trader enters the market at time zero with x_0 dollars in bonds and y_0 dollars in *ex dividend* shares of stock, but (in addition to the first scenario) immediately writes an American futures call option with maturity $T, T \leq T^F$, where C are the net cash proceeds from writing the call.[4] We assume that the trader may not trade the call option thereafter.[5] At each trading date t ($t = 0, 1\ldots, T$) the trader is informed whether or not she has been assigned (that is, assigned to act as the counterparty of the holder of a call who exercises the call at that time). If the trader has been assigned, the call position is closed out, the trader pays $F_t - K$ in cash, and the value of the cash account decreases from x_t to $x_t - (F_t - K)$. The trader makes sequential investment decisions with the objective to maximize expected utility, $E[u_{T'}(W_{T'})]$. We refer to this trader as the *option* (plus index and bond) trader, OT, and denote her maximized expected utility by $V_0^{OT}(x_0 + C, y_0)$.

---

[3] Alternatively, the objective may be the maximization of the discounted sum of the utility of consumption $u_t(c_t)$ at each trading date, including the terminal date. In this case, the terminal date may be finite or infinite. Although the Constantinides and Perrakis (2007) bounds are derived under the terminal wealth objective, they remain valid without any reformulation under the alternative objective.

[4] The reservation write price of a call is derived from the perspective of a trader who is marginal in the index, the bond, and only one type of call or put option at a time. Therefore, these bounds allow for the possibility that the options market is segmented.

[5] The reservation write price of a call is derived under this constrained policy. Under this policy, the investor increases her expected utility by writing a call at price $\overline{C}$ and refraining from trading the call thereafter. If the constraint on trading the call is relaxed, the policy which the investor follows under the constraint policy remains feasible and increases her expected utility by writing a call at price $\overline{C}$. Therefore, $\overline{C}$ remains an upper bound on the reservation write price of a call. Whereas the upper bound may be tightened when the constraint on trading the call is relaxed, there is no known tighter bound that is preference free. For further discussion on this point, see Constantinides and Perrakis (2007).

For a given pair $\left(x_0, y_0\right)$, we define the reservation write price of a call as the value of C such that $V_0^{OT}\left(x_0 + C, y_0\right) = V_0^{IT}\left(x_0, y_0\right)$. The interpretation of C is the write price of the call at which the trader with initial endowment $\left(x_0, y_0\right)$ is indifferent between writing the call or not. Constantinides and Perrakis (2007) state a tight upper bound on the reservation write price of a European futures call option that is independent of the trader's utility function and initial endowment and independent of the early exercise policy on the calls:

$$\overline{C}\left(F_t, S_t, t\right) = \frac{1+k_1}{1-k_2}\max\left[N\left(S_t, t\right),\ F_t - K\right], \qquad t \le T. \tag{1}$$

The function $N(S,t)$ is defined as follows:

$$\begin{aligned} N(S,t) &= (R_S)^{-1} E[\max\{\left(1+\gamma\right)^{-\left(T^F - t - 1\right)} R^{T^F - t - 1} S_{t+1} + \overline{\varepsilon} - K,\, N(S_{t+1}, t+1)\} \big| S_t = S], \quad t \le T-1 \\ &= 0, \quad t = T. \end{aligned} \tag{2}$$

The economic interpretation of the call upper bound is as follows. If we observe a call bid price above the reservation write price, $\overline{C}$, then any trader (as defined in this paper) can increase her expected utility by writing the call.

Transaction costs on the index have only a small effect on the upper bound. Specifically, without transaction costs on the index, the upper bound is $\max\left[N\left(S_t, t\right),\ F_t - K\right]$; with transaction costs on the index, the upper bound merely increases by the multiplicative factor $\left(1+k_1\right)/\left(1-k_2\right)$. The explanation is that this particular bound is based on a comparison of the utility of an index trader and the utility of an option trader. Both traders follow the trading policy which is optimal for the index trader but is generally suboptimal for the option trader. This policy incurs very low transaction costs because the trader trades infrequently, as shown in Constantinides (1986).

If we further assume that the trader can buy a call at price $\overline{C}(F_t, S_t, t)$ or less and trade the futures and do so costlessly, we obtain the following put upper bound:[6]

$$\overline{P}(F_t, S_t, t) = \overline{C}(F_t, S_t, t) - R^{-(T-t)} F_t + K, \quad t \le T. \tag{3}$$

The interpretation of the put upper bound is as follows. If we observe a put bid price above the reservation write price $\overline{P}$, then any trader can increase her expected utility by writing the put.

Constantinides and Perrakis (2007) also stated a tight lower bound on the reservation purchase price of an American futures put option. The cash payoff of the put exercised at time t is $K - F_t,\ t \le T$. As in the case of a call option, we define the reservation purchase price of a put as the value of P such that the trader with initial endowment (x_0, y_0) is indifferent between purchasing the put or not. The following is a tight lower bound on the reservation purchase price of an American futures put option that is independent of the trader's utility function and initial endowment:

$$\underline{P}(F_t, S_t, t) \equiv \max\left[K - F_t,\ \frac{1-k}{1+k} M(S_t,\ t)\right], \quad t \le T. \tag{4}$$

The function $M(S, t)$ is defined as follows:

$$\begin{aligned} M(S,t) &= (R_S)^{-1} E\left[\max\left[K - (1+\gamma)^{-(T^F - t - 1)} R^{T^F - t - 1} S_{t+1} - \overline{\varepsilon},\ M(S_{t+1}, t+1)\right] \mid S_t = S\right], \quad t \le T-1 \\ &= 0, \quad t = T. \end{aligned} \tag{5}$$

[6] We prove equation (3) by noting that an investor achieves an arbitrage profit by buying a call at $\overline{C}(F_t, S_t, t)$, writing a put at $P, P > \overline{P}(F_t, S_t, t)$, selling one futures, and lending $K - R^{-(T-t)} F_t$. In the proof, we ignore the daily marking-to-market on the futures until the exercise of the put or the options' maturity, whichever comes first.

If we observe a put ask price below the reservation purchase price $\underline{P}$, then any trader can increase her expected utility by buying the put. As in the case of the upper call bound, transaction costs on the index have only a small effect on the lower put bound.

If we further assume that the trader can write a put at price $\underline{P}(F_t, S_t, t)$ or more, and trade the futures and do so costlessly, then we obtain the following call lower bound, with corresponding interpretation:[7]

$$\underline{C}\left(F_t, S_t, t\right) = \underline{P}\left(F_t, S_t, t\right) + R^{-(T-t)} F_t - K, \quad t \leq T. \tag{6}$$

If we observe a call ask price below the reservation purchase price $\underline{C}$, then any trader can increase her expected utility by buying the call.

2 Empirical Design

We describe our empirical design, starting with a description of the data, the calibration of a tree of the daily index return, and the construction of the portfolio of the index trader (who does not trade in the option) and of the option trader. This allows us to introduce the well-known Sharpe ratio test and we discuss the problems associated with using this test. To address problems with the Sharpe ratio test, we introduce tests based on second order stochastic dominance.

2.1 Data and estimation

We obtain the time-stamped quotes of the 30-calendar-day S&P 500 futures options and the underlying 1-month futures for the period February 1983-July 2006 from the Chicago Mercantile Exchange (CME) tapes. This results in 247 sampling dates. We obtain the interest rate as the three-month T-bill rate from the Federal Reserve Statistical Release. The data sources are described in further detail in Appendix A.

[7] We prove equation (6) by noting that an investor achieves an arbitrage profit by writing a put at $\underline{P}\left(F_t, S_t, t\right)$, buying a call at $C, C < \underline{C}\left(F_t, S_t, t\right)$, selling one futures, and lending $K - R^{-(T-t)} F_t$.

For the daily index return distribution, we use the historical sample of log returns from January 1928 to January 1983. However, when looking forward for each of our 247 option sampling dates, we adjust the first four moments of the index return distribution in various ways which we now describe in detail. We set the mean index return at 4% plus the observed 3-month T-bill rate instead of estimating the mean index return from the data in order to mitigate statistical problems in estimating the mean.[8] We implement this by adding a constant to the observed logarithmic index returns so that their sample mean equals the above target. We estimate the 3rd and 4th moments of the index return as their sample counterparts over the preceding 90 days.

Finally, we estimate both the unconditional and conditional volatility of the index return as follows. We estimate the unconditional volatility as the sample standard deviation over the period January 1928 to January 1983.[9] We estimate the conditional volatility in three different ways: (1) the sample standard deviation over the preceding 90 trading days;[10] (2) the at-the money (ATM) implied volatility (IV) on the preceding day, adjusted by the mean prediction error for all dates preceding the given date (typically some 3%), where we drop from the preceding days all 21 pre-crash observations; and (3) the GARCH volatility using GARCH coefficients estimated for S&P 500 daily returns over January 1928 to January 1983 applied to residuals observed over the 90 days preceding each sample date to form projections of the volatility realized till the option expiry date. In Table 1, we report statistics of the prediction error of the above volatility estimates. The best overall predictors are the adjusted ATM IV and the 90-day historical volatility.

2.2 Calibration of the index return tree and calculation of the option bounds

We model the path of the daily index return till the option expiration on a T-step tree, where T is the number of trading days in that particular month.[11] The tree is recombining with m

---

[8] Short-horizon forecasts of the *conditional* mean equity premium are notoriously unreliable. Fama and French (2002), Constantinides (2002), and Dimson, Marsh and Staunton (2006) estimated the adjusted *unconditional* mean equity premium to be 4-6% per year. For our main results, we set the mean return at 4% plus the observed 3-month T-bill rate. We also report results when we set the mean return at 6% plus the observed 3-month T-bill rate.

[9] We have also estimated the unconditional volatility over the 24 years prior to January 1983. The results remain essentially unchanged and are not reported in the paper.

[10] We have also estimated the conditional volatility over the preceding 360 days. The results remain essentially unchanged and are not reported in the paper.

[11] For example, if the 3rd Friday of July is on July 27, we record the price of the July option on June 27,

branches emanating from each node. Every month we calibrate the tree by choosing the number of branches and the return at each node to match the first four moments of the daily index return distribution, as described in Appendix B.

The upper and lower bounds on the call and put prices are given in equations (1)-(6). We numerically calculate the bounds by iterating backwards on the calibrated tree.

2.3 Portfolio construction and trading

For each monthly stock return path, we employ the following trading policies. For the index trader (who manages a portfolio of the index and the risk free asset in the presence of transaction costs), we employ the optimal trading policy, as derived in Constantinides (1986) and extended in Perrakis and Czerwonko (2007) to allow for dividend yield on the stock. Essentially, this policy consists of trading only to confine the ratio of the index value to the bond value, y_t / x_t, within a no-transactions region, defined by lower and upper boundaries. We derive these boundaries for the following parameter values: one-way transaction cost rate on the index of 0.5%; annual return volatility of the index of 0.1856, the sample volatility over 1928-1983; interest rate equal to the observed 3-month T-bill date; risk premium 4%; and constant relative risk aversion coefficient of 2.[12] For this set of parameters, the lower and upper boundaries are $y_0 / x_0 = 1.2026$ and 1.5259, respectively. At the beginning of each month and before the trader trades in options, we set $x_0 = 73{,}300$ and $y_0 = 100{,}000$, which corresponds to the midpoint of the no-transactions region, $y_0 / x_0 = 1.3642$.

For the option trader (who manages a portfolio of the option, index, and the risk free asset in the presence of transaction costs), we employ the trading policy which is optimal for the index trader but is generally suboptimal for the option trader. Recall that the goal is to demonstrate that there exist profitable investment opportunities for the option trader. Given this goal, it suffices to show that there exist profitable investment opportunities for the

which is 30 calendar days earlier. (If June 27 is a holiday, we record the price on June 26.) If there are 21 trading days between June 27 and July 27, we model the path of the daily index return till the option expiration on a 21-step tree.

[12] We clarify that the upper and lower stochastic dominance bounds on option prices apply to any risk averse trader, independent of her particular degree of risk aversion. In our empirical work, we make an assumption about the relative risk aversion coefficient in order to calculate the boundaries of the no-

option trader even if the option trader follows a generally suboptimal policy. We set x_0 and y_0 to the same values as for the index trader. However, this portfolio composition changes, depending on the assumed position in futures options, as explained in Appendix C.

We focus on the cases where the basis risk bound, $\bar{\varepsilon}$, is 0.5% of the index price. Over the years 1990-2002, 95% of all observations have basis risk less than 0.5% of the index price. For reference purposes, we also consider the case $\bar{\varepsilon} = 0$. As to be expected, when we suppress the basis risk, the bounds are tighter and there appear to be more violations.

2.4 Description of the empirical tests

For each one of our methods of estimating the bounds, we obtain 247 monthly portfolio returns for the index trader and the option trader, respectively. Our goal is to test whether the portfolio profitability of the index and option traders are statistically different in the months in which we observe violations of the bounds.

In our first set of tests, we compare the Sharpe ratios of the two portfolios. Despite the well-known limitations of the Sharpe ratio, we report these results because the Sharpe ratio is one of the most popular measures of portfolio performance.[13] We use the approach of Jobson and Korkie (1981) with the Memmel (2003) correction that accounts for different variances of the two portfolios. Details of the test are described in Appendix D.

In our second set of tests, we compare the returns of the two portfolios in terms of the criterion of stochastic dominance, which states that the dominating portfolio is preferred by any risk-averse trader, independent of distributional assumptions such as normality and preference assumptions such as quadratic utility. Specifically, we test the null hypothesis $H_0 : OT \not\succ_2 IT$, which states the option trader's portfolio return does not stochastically dominate the index trader's portfolio return, against the alternative hypothesis $H_A : OT \succ_2 IT$, which states the option trader's portfolio return stochastically dominates the

---

transactions region for a specific trader. We present results for relative risk aversion 2 and 10.

[13] The Sharpe ratio ignores moments of the return distribution beyond the mean and variance and this is theoretically justified only the special cases where either investors have quadratic utility or the portfolio returns are normally distributed. The latter assumption is obviously violated in portfolios that include options.

index trader's portfolio return. We report the results of tests proposed by Davidson and Duclos (2006), using the algorithm developed by Davidson (2007).

An earlier test, proposed by Davidson and Duclos (2000), tests the null hypothesis $H_0 : OT \succ_2 IT$ against the alternative, which is that either $IT \succ_2 OT$ or that neither one of the two distributions dominates the other. Hence, rejection of the null hypothesis fails to rank the two distributions in the absence of information on the power of the test, which is generally not available. We report results of this test as well because it has certain statistical advantages over the Davidson and Duclos (2006) test. Appendix D provides details for both tests.

3 Empirical Results

In Section 3.1, we describe the empirical results. We compare the portfolio return of an option trader who writes overpriced calls, puts, or straddles at their bid price with the portfolio return of an index trader who does not trade in the options over the period 1983-2006. In out-of-sample tests, we find that the return of an option writer stochastically dominates the index trader's return, net of transaction costs and the bid-ask spread. We also find that the Sharpe ratio of the option trader's return is higher than the Sharpe ratio of the index trader's return and the difference is often statistically significant. In Section 3.2, we establish that the empirical results are robust. In Section 3.3, we demonstrate that trading policies triggered by violations of the stochastic dominance bounds consistently outperform naïve filter rules of buying low and selling high.

3.1 Results

In Figure 1, we plot the four bounds for one-month options, expressed in terms of the implied volatility, as a function of the moneyness, K / F_0. We set $\sigma = 20\%$ and $\bar{\varepsilon} = 0$. The figure also displays the 95% confidence interval, derived by bootstrapping the 90-day distribution. Regarding the upper bounds, we observe that the call upper bound is tighter than the put upper bound. Also, the call and put upper bounds are tighter when the (*K/F*) ratio is high, that is when the calls are OTM or the puts are ITM. Regarding the lower bounds, we observe that the put lower bound is tighter than the call lower bound. Also, the call and put lower bounds are tighter when the (*K/F*) ratio is low, that is when the calls are ITM or the puts are OTM.

In Figure 2, we display the time pattern of actual violations of the call upper bound. For all different ways of estimating volatility, we observe violations after significant down moves in the index, i.e., when we expect the implied volatility of traded options to be high. We do not present time patterns for the remaining three bounds since we do not observe many violations of these bounds.

In Table 2, we present the cases of call and put bid prices violating their upper bound, when we set the basis risk bound at 0.5% of the index price. We do not present the cases of call and put ask prices violating their lower bound because we do not have a sufficient number of such violations to be able to draw statistical inference, as we observed in Figure 2. We find a higher frequency of violations of the upper call bound than of the upper put bound since the upper call bound is tighter than the upper put bound, as we observed in Figure 1.

The Sharpe ratio of the call trader's return is uniformly higher than the Sharpe ratio of the index trader's return, irrespective of the mode of predicting the volatility as an input to the call upper bound. When the call upper bound is calculated using the adjusted IV or the GARCH volatility, the difference in Sharpe ratios exceeds 9% annually and is statistically significant at the 10% level. There are far fewer violations of the put upper bound and, therefore, the results are statistically weaker. Nevertheless, when using the unconditional prediction of volatility as an input to the put upper bound, we find 23 violations of the put upper bound and the put trader's portfolio has a Sharpe ratio that exceeds the index trader's portfolio by 12.8%, statistically significant at the 10% level. These Sharpe ratio preliminary results motivate and reinforce our main results on stochastic dominance which are discussed next.

The DD (2000) test does not reject the hypothesis $H_0: OT \succ_2 IT$, which states that the option trader's return dominates the index trader's return; and rejects the hypothesis $H_0: IT \succ_2 OT$, which states that the index trader's return dominates the option trader's return. Thus, we unfortunately cannot decide between dominance of the OT strategy or a tie where we cannot establish dominance one way or another. Luckily, the DD (2006) test allows us to make stronger claims of dominance. The DD (2006) test strongly rejects the null hypothesis $H_0: OT \not\succ_2 IT$, which states that either the index trader's return dominates the option trader's return or that neither distribution dominates the other. The p-values of the hypothesis $H_0: IT \not\succ_2 OT$ are equal to one and are not reported here. Here, we can uniquely establish dominance of the OT strategy.

Next, we explore the performance of the policy of writing overpriced calls through the policy of writing straddles. Straddles are popular trading policies and have been previously investigated in the literature. For example, Coval and Shumway (2001) show that a long ATM straddle on the S&P 500 index or the S&P 100 index produces substantial negative returns.[14] Each month, we look for call bid prices that lie above the upper call bound. If we find at least one call bid prices that lie above the upper call bound *and* if we find at least one put bid price (irrespective of whether the put bid price violates the put upper bound or not) we proceed as follows. We short equal fractions of the calls that violate the call upper bound, such that the fractions add up to one; we short equal fractions of the puts for which we have bid prices, such that the fractions add up to one; and we sell one futures on the index. The results are reported in Table 3, panel A. The annualized Sharpe ratio differentials are large and significant at the 5% or 1% level. These results are consistent with the results of Coval and Shumway (2001). The DD (2000) test does not reject the hypothesis $H_0: OT \succ_2 IT$. It often rejects the hypothesis $H_0: IT \succ_2 OT$, but not consistently so. Finally, the DD (2006) test strongly rejects the hypothesis $H_0: OT \not\succ_2 IT$. We conclude that the results in Table 3, panel A, are consistent with those in Table 2.

We note that the numbers of cross-sections for which short straddles are traded in Table 3, panel A, is significantly lower than the corresponding numbers for calls is Table 2. Since, in our approach, the straddle sales are solely determined by the violations of the call upper bound, we increase the number of cross-sections by relaxing the requirement that the put sale has to occur at the same strike price as that of the call that triggers the violation. Instead, we require that the moneyness of the put remains within 0.98-1.02 times the moneyness of the triggering call. The first put quote within this bound following the call

[14] The paper of Coval and Shumway (2001)) focuses on the relation between the CAPM beta and the return of straddles and, as such, differs from our goal of measuring the performance of straddles, net of bid-ask spreads and through the broader criterion of stochastic dominance.

violations is included in the ensuing straddle position. The results for this approach are presented in Table 3, panel B. Compared to Table 3, panel A, we observe an improvement in the stochastic dominance tests results and a systematic increase in the Sharpe ratios.

Since the results for straddles imply that the call upper bound is an efficient selector of overpriced puts, we apply this selection criterion to all available put quotes. Specifically, for every put bid we derive the call upper bound. Then we sell the put if its implied volatility exceeds the corresponding quantity for the call upper bound. The results are reported in Table 2, panel C. The returns of the put selling policy stochastically dominate the returns of the index trader's portfolio, irrespective of the way in which volatility is estimated. The put selling policies produce Sharpe ratios that are higher than the Sharpe ratios of the index trader's portfolio returns, irrespective of the way in which the volatility is estimated; the Sharpe ratio differences are statistically significant when the volatility is estimated as the adjusted IV or as the GARCH volatility.

3.2 Robustness tests

In Tables 4-9, we demonstrate that the results of Tables 2 and 3 are robust. Table 4 differs from Table 2 only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the basis risk by $\bar{\varepsilon} = 0.5$. There are now more options across the board violating the bounds because all the bounds become tighter: the upper bounds are lowered and the lower bounds are raised. We present the cases of call and put bid prices violating their upper bound. We do not present results for the cases when the call and put ask prices violate their lower bound because we still do not have a sufficient number of such violations to be able to make statistical inference.

Since the upper call and put bounds are lower, the options trader is less selective than before in writing options that violate their upper bounds and we find that the differences of the Sharpe ratios are smaller in Table 4 than in Table 2. However, since there are more observations in Table 4, the differences of the Sharpe ratios are statistically more significant than in Table 2. The DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$ and rejects the hypothesis $H_0 : IT \succ_2 OT$. Finally, the DD (2006) test strongly rejects the hypothesis $H_0 : OT \not\succ_2 IT$. We conclude that the results in Table 3 are consistent with those in Table 2.

Table 5 differs from Table 3 on straddles only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the basis risk by $\bar{\varepsilon} = 0.5$. Again, we conclude that the results in Table 5 are consistent with those in Table 3.

Table 6 differs from Table 2 only in that the relative risk aversion coefficient is set at 10 instead of 2. Since the upper and lower stochastic dominance bounds on option prices are independent of the trader's utility, we observe the same number of violations in Table 6 as we do in Table 2. The change in the risk aversion coefficient does change the boundaries of the no-transactions region and, therefore, the trading policy of the index trader and the option trader. The Sharpe ratio differences are substantially higher in Table 6 but these differences are not statistically significant. (Recall that the differences in Panels A and B of Table 2 are only marginally significant). The stochastic dominance results in writing calls are as strong in writing calls and stronger in writing puts.

Table 7 differs from Table 2 only in that the expected premium on the index is set at 6% instead of 4%. The differences in Sharpe ratios are comparable to those in Table 2 but these differences are not statistically significant. The stochastic dominance results in writing calls are as strong in writing calls and stronger in writing puts. We conclude that the results in Table 2 are robust to the assumption that the expected premium on the index is 4%.

Table 8 differs from Table 2 only in that we exclude from the sample the seven month from October 1987 to April 1988 in order to abstract from effects associated with the crash. The difference in Sharpe ratios between the returns of the option trader and index trader are comparable to those in Table 2, but are not statistically significant. The stochastic dominance results in writing calls and puts are the same as in Table 2. This is partly due to the fact that stock prices recovered on the days following the crash and the October return on the index is flat.

The bounds that are used in identifying mispriced options in our empirical work are calculated with parameter inputs which are point estimates and vary for each time point of our sample for all but the historical method of estimating the bounds. These varying parameters imply that the screening rules for mispriced options become conditional on the time point of our sample. Since the earlier tests do not recognize this conditionality, we develop in Appendix E an alternative set of tests that explicitly take into account the time varying nature of our sample and conclude that conditional and unconditional tests lead to same conclusions. The results are reported in Table 9 and discussed in Appendix E. They are also consistent with the main results of Table 2 and supportive of the mispricing hypothesis, even though they are derived with a different method.

3.3 Comparison with naïve trading

The option trading policies triggered by violations of the stochastic dominance bounds superficially resemble a well-known naïve trading policy of buying (or selling) an option when its IV is at the low (or, high) end of the IV distribution of options within a certain range of moneyness. In this section, we demonstrate that a particular form of this naïve trading policy consistently underperforms the earlier trading policies triggered by violations of the stochastic dominance bounds.

We derive the 90th, 97.5th, 10th, and 2.5th percentiles of the IV distribution for a given range of moneyness by applying the method of Yu and Jones (1998).[15] Table 10 presents the results for the naïve trading policy. In all trading policies, we mirrored the policies applied for the option trader (OT). We observe that the number of cross-sections for which we find 'quantile violations' is relatively low. This observation is caused by the clustering of violations in some cross-sections. We conclude that the naïve trading policy detects parallel shifts in the implied volatility instead of singling out unusual observations in the majority of cross-sections. These shifts appear to be inefficient: the naïve bounds only capture some of these parallel shifts in implied volatility, namely violations at the top. The naïve trading policy performs well on the sell side but performs disastrously on the buy side, as shown by the stochastic dominance statistics and Sharpe ratios.

---

[15] The quantile regression is a kernel regression in two dimensions, in our case in the dimensions of moneyness and IV. As is usual in a kernel regression, the critical part is in determining the kernel bandwidth. To determine this quantity in the moneyness dimension, we use the Leave-One-Out method, as described in Härdle (1990), for which we use the transformation given in Table 1 in Yu and Jones (1998). To determine the bandwidth in the implied volatility dimension, we use (12) in Yu and Jones (1998). As our sample to derive the critical quantile function, we use five past observations with 30 days to maturity. Using ten past observations yields similar results, not reported here. We verified that in-sample the likelihood of observations outside any critical quantile q is close to $\min\left(q, |1-q|\right)$.

4 Concluding Remarks

We search for mispriced American call and put options on the S&P 500 index futures by employing stochastic dominance upper and lower bounds on the prices of options. We identify call and put bid prices on index futures that violate the upper bounds and call and put ask prices that violate the lower bounds. We find a substantial number of violations of the upper bounds, but relatively few violations of the lower bounds. Since the frequency of violations of the lower bounds is too low for statistical inference, we focus on violations of the upper bounds.

We compare the portfolio return of an option trader who writes overpriced calls or puts at their bid price with the portfolio return of an index trader who does not trade in the options over the period 1983-2006. In out-of-sample tests, our main result is that the return of a call or put writer stochastically dominates (in second order) the index trader's return, net of transaction costs and the bid-ask spread. The dominance holds under a variety of methods in estimating the underlying return distribution. It also holds with or without the assumption that the portfolio returns are drawn from the same distribution each period.

We also find that the Sharpe ratio of the call trader's return is uniformly higher than the Sharpe ratio of the index trader's return and is often statistically significant. The Sharpe ratio of the put trader's return is uniformly higher than the Sharpe ratio of the index trader's return but the results are less statistically significant. Finally, the policy of writing straddles produces returns that strongly stochastically dominate the index trader's return and have substantially higher Sharpe ratios. The results are supportive of the hypothesis that the options identified by violations of the CP (2007) bounds are mispriced.

Appendix A: Data

We obtain the time-stamped quotes of the one-month S&P 500 futures options and the underlying one-month futures for the period February 1983-July 2006 from the CME tapes. From the futures prices, we calculate the implied S&P 500 index prices by applying the cost-of-carry relation $F_t = (1+\gamma)^{-(T^F - t)} R^{T^F - t} S_t + \varepsilon_t$, assuming away basis risk, $\varepsilon_t \equiv 0$.[16] We obtain the daily dividend record of the S&P 500 index over the period 1928-2006 from the S&P 500 Information Bulletin and convert it to a constant dividend yield for each 30-day period. Before April 1982, dividends are estimated from monthly dividend yields. We obtain the interest rate as the three-month T-bill rate from the Federal Reserve Statistical Release. We estimate the variance of the basis risk, $\mathrm{var}(\varepsilon_t)$, from the observed futures prices and the intraday time-stamped S&P 500 record obtained from the CME.

We rescale the index price S_t by the multiplicative factor $100,000/S_0$ so that the index price at the beginning of each 30-day period is 100,000. Accordingly, we rescale the futures price, index futures option price, and strike by the same multiplicative factor.

We consider options maturing in 30 calendar days, which results in 247 sampling dates.[17] Since the first maturity of serial options was in August 1987, the first 19 periods occur with quarterly periodicity. Overall, we record 36,921 raw call quotes and 42,881 raw put quotes. After eliminating obvious data errors, we apply the following filters: minimum 15 cents for a bid quote and 25 cents for an ask quote; *K/F* ratio within 0.96-1.08 for calls and within 0.92-1.04 for puts; and matching the underlying futures quote within 15 seconds. Part of the data is lost due to the CME rule of flagging quotes, i.e. bids (asks) are flagged only if a bid (ask) is higher (lower) than the preceding bid (ask); in addition, no transaction data is flagged. We recover a large part of the data by analyzing the sequence between consecutive bid-ask flags; however, this recovery is not possible in all cases. As a result of the applied filters, we obtain 29,822 quotes for calls and 30,281 quotes for puts in our final

[16] Recall that our goal is to compare the investment policies, of the index trader and the option trader. Since both policies stipulate approximately the same stock component, the effect of this component cancels each other out. Also, it is a common empirical approach to derive the index value from the index futures; see, for example, Jackwerth and Rubinstein (1996).

[17] The 30-day rule eliminates the occurrence of the October crash from our sample. Therefore, we use one 40-day period to have the crash (the 248th observation) and verified that the inclusion of the crash does not alter our results.

sample. These quantities translate into roughly 60 data points for all strikes for either bid or ask prices for an average day.

Appendix B: Calibration of the index return tree

We model the paths of the daily index return on a recombining tree with *m* branches emanating from each node. Every month we calibrate the tree, including *m*, to match the 2nd, 3rd and 4th moments of the daily index return distribution. For the 3rd and 4th moments, we always use the observed sample moments over the 90 preceding calendar days.

In the first step of our algorithm, we pick a value for the number of branches *m* and group the sample of daily returns in a histogram with *m* bins of equal length (on the log scale) such that the extreme bins are centered on the extreme observed returns. The center of each bin then becomes a state in the equally spaced lattice, with the ordered states and the corresponding probabilities denoted respectively as x_i and p_i, $i = 1...m$. In the second step, we impose the desired first three moments by altering the lattice from the step one. We derive the adjustments by solving the following set of three non-linear equations that are simply three moment conditions for the quantities *a*, *b,* and *c*:

$$\sum_{i=1}^{m} p_i^* \exp\left(ax_i + b\right) - \exp\left(\hat{\mu}\right) = 0$$

$$\sum_{i=1}^{m} p_i^* \left[\exp\left(ax_i + b\right)\right]^2 - \exp\left(\hat{\mu}\right)^2 - \hat{\sigma}^2 = 0 \qquad \text{(B.1)}$$

$$\sum_{i=1}^{m} p_i^* \left[\exp\left(ax_i + b\right) - \exp\left(\hat{\mu}\right)\right]^3 - \hat{\mu}_3 \hat{\sigma}^3 = 0$$

where $\exp\left(\hat{\mu}\right)$ and $\hat{\sigma}^2$ are the first and second target moments, respectively, $\hat{\mu}_3$ is the sample skewness, and $p_i^* \equiv \dfrac{p_i + c\underline{1}_{\left(i \ge n^*\right)} \underline{1}_{\left(p_i \ne 0\right)}}{\sum_{i=1}^{m}\left(p_i + c\underline{1}_{\left(i \ge n^*\right)} \underline{1}_{\left(p_i \ne 0\right)}\right)}$, where $\underline{1}_{(.)}$ is the indicator function, n^* is the index to this x_i which brackets from above the target expected log-return $\hat{\mu}$. The first indicator function ensures that the constant c is added only to the probabilities in the right tail of the distribution; the second one ensures that the constant c is added only to the positive probabilities.[18,19] Note that the affine transformation of the log-states x_i preserves the equal distance between the adjacent states, which is necessary for the lattice to recombine.

To match the fourth sample moment $\hat{\mu}_4$, we resort to varying m, the number of nodes in the lattice. With each new m the initial distribution derived from a histogram changes providing some variability in the fourth moment after the adjustments resulting from solving (B.1). After a search over a range of m's, we pick this distribution which has the lowest absolute difference between its kurtosis and the sample kurtosis $\hat{\mu}_4$. It turns out that this search procedure ends up with acceptably small errors in matching $\hat{\mu}_4$ for the data that we use. For the four volatility prediction modes we apply in our work, the relative error on the fourth moment had the following characteristics: median 0.003%, 99th percentile 0.105%, maximum 1.659% across 973 observations while we constrained the lattice size m to be no larger than 201.[20]

Instead of using a histogram in the first step above, we could start building our lattice by discretizing a kernel-smoothed distribution. However, since the kernel smoothing would involve more parameter choices and would result in a significantly larger lattice size to attain accuracy similar to the one of our method, we retain a preference for the histogram approximation.[21]

---

[18] By grouping the observation in the histogram as a rule we end up with states in our lattice that have zero probabilities. We don't investigate here the precise recombination pattern of a lattice with zero-probability states; we observe, however, that the number of zero-probability states remains relatively constant as the number of convolutions of the lattice with itself increases, resulting in a decreasing proportion of such states as the time period increases.

[19] Note that the presented adjustment of the probabilities in the right tail may not yield an admissible solution, i.e. we may end up with some negative probabilities. If this is the case, we introduce an analogous adjustment in the left tail of the distribution.

[20] This lattice size appears unattractive to derive recursive conditional expectations. However, the use of fast Fourier transforms results in a fairly short processing time. See Cerny (2004).

[21] A critical parameter in the kernel density estimation is the kernel bandwidth. In addition, since the density estimate of the log-returns covers the real line, the scope of the discretized distribution would need to be chosen.

Appendix C: Trading policy

We consider calls with moneyness (*K/S*) within the range 0.96-1.08 and puts within the range 0.92-1.04. If we observe *n* call bid prices violating the call upper bound, each with different strike price, the option trader writes $1/n$ calls of each type with the underlying futures corresponding to the index value of y_0. The trader transfers the proceeds to the bond account: $x = x_0 + \sum_{i=1}^{n} C_i / n$ and $y = y_0$.

If we observe *n* put ask prices violating the put lower bound, each with different strike price, the option trader buys $1/n$ puts of each type and finances the purchase out of the bond account: $x = x_0 - \sum_{i=1}^{n} P_i / n$ and $y = y_0$.

However, when there is a violation of the upper put bound and the option trader writes puts, the trader also sells one futures contract for each written put. The intuition for this policy may be gleaned from the observation that the combination of a written put and a short futures amounts to a synthetic short call. In fact, the upper put bound in equation (3) is derived from the upper call bound in equation (2) through the observation that if we can write a put at a sufficiently high price we violate the upper call bound by writing a synthetic call.[22]

Finally, when there is a violation of the lower call bound and the option trader buys calls, the trader also sells one futures contract for each purchased call. The intuition is the same as above.

The early exercise policy of a call is based on the function *N* in equation (2). The early exercise policy of a put is based on the function *M* in equation (5). However, whenever the option trader is short an option, each period we derive the functions *N* and *M* based on the *forward-looking* distribution of daily returns, i.e. these functions are derived under the empirical distribution of the daily index returns between the option trade and the option maturity. Effectively, we endow the counterparty of the option trader with information on the 2nd, 3rd, and 4th moments of the forward distribution, while imposing the

[22] In implementing the trading policy of either writing puts or buying calls, the option trader buys or sells a futures contract as well and this violates the assumption made in Section 1 that the option trader does not trade in futures. Even when we relax the assumption on trading in futures, in practice, traders manage their portfolio by trading in the index because of the inconvenience and cost of the frequent rolling over of short-term futures contracts and the illiquidity of long-term futures and forward contracts.

first moment. The early exercise policy of a call or put is simplified by the observation that the decision is a function only of time and the ratio of the strike price to the index level.

Appendix D: The Sharpe ratio and the Davidson-Duclos (2000, 2006) tests

For the Sharpe ratio tests, we use the approach of Jobson and Korkie (1981) with the Memmel (2003) correction that accounts for different variances of the two portfolios. Specifically, given the sample of *N* realizations of the index trader's (IT) and option trader's (OT) portfolio outcomes with $\hat{\mu}_{OT}$, $\hat{\mu}_{IT}$, $\hat{\sigma}^2_{OT}$, $\hat{\sigma}^2_{IT}$, $\hat{\sigma}_{IT,OT}$ as their estimated excess means, variances, and covariances, we test the hypothesis $H_0 : \hat{\mu}_{OT}\hat{\sigma}_{IT} - \hat{\mu}_{IT}\hat{\sigma}_{OT} \leq 0$ with the test statistic $\hat{z}$, which is asymptotically standard normal:

$$\hat{z} = \frac{\hat{\mu}_{OT}\hat{\sigma}_{IT} - \hat{\mu}_{IT}\hat{\sigma}_{OT}}{\sqrt{\hat{\theta}}} \tag{D.1}$$

where

$$\hat{\theta} = \frac{1}{N}\left(2\hat{\sigma}^2_{IT}\hat{\sigma}^2_{OT} - 2\hat{\sigma}_{IT}\hat{\sigma}_{OT}\hat{\sigma}_{IT,OT} + \frac{1}{2}\hat{\mu}^2_{OT}\hat{\sigma}^2_{IT} + \frac{1}{2}\hat{\mu}^2_{IT}\hat{\sigma}^2_{OT} - \frac{\hat{\mu}_{IT}\hat{\mu}_{OT}}{\hat{\sigma}_{IT}\hat{\sigma}_{OT}}\hat{\sigma}^2_{IT,OT} \right). \tag{D.2}$$

DD (2000) provide a test of the null hypothesis $H_0 : OT \succ_2 IT$ in terms of the maximal and minimal values of the extremal test statistic, $\hat{T}(z)$. The null is not rejected, if the maximal value of the statistic is positive and statistically significant *and* the minimal value of the statistic is either positive or negative and statistically not significant.

The variable *z* denotes the logarithm of end-of-the-month wealth of a trader, where the subscripts *IT* and *OT* distinguish between the index trader and the option trader. The statistic $\hat{T}(z)$ is defined as follows:

$$\hat{T}(z) = \frac{\hat{D}^2_{IT}(z) - \hat{D}^2_{OT}(z)}{\sqrt{\hat{V}^2(z)}} \tag{D.3}$$

where

$$\hat{D}_I^2(z) = \frac{1}{N}\sum_{i=1}^{N}\left(z - W_{Ji}\right)_+ \tag{D.4}$$

$$\hat{V}^2(z) = \hat{V}_{IT}^2(z) + \hat{V}_{OT}^2(z) - 2\hat{V}_{IT,OT}^2(z) \tag{D.5}$$

$$\hat{V}_I^2(z) = \frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(z - W_{Ii}\right)_+^2 - \hat{D}_I^2(z)^2\right],\ \ I = IT, OT \tag{D.6}$$

and

$$\hat{V}_{OT,IT}^2(z) = \frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(z - W_{ITi}\right)_+\left(z - W_{OTi}\right)_+ - \hat{D}_{IT}^2(z)\hat{D}_{OT}^2(z)\right]. \tag{D.7}$$

The maximal and minimal values of the statistic are calculated as a maximum and minimum of (D.3) over a set of points of z as explained below. Stoline and Ury (1979) provide tables for the distribution of the maximal and minimal value of $\hat{T}(z)$, which is not standard at the levels 1, 5 and, 10%. In principle, the number of points in this joint support over which the test may be performed needs to be restricted since a 'large' number of these points violates the independence assumption between the $\hat{T}(z)$s,. Therefore, we compute these statistics for 20 points equally spaced in the log-transformed joint support of W_{IT} and W_{OT}, which corresponds to k = 20 in the Stoline and Ury (1979) tables.

DD (2006) provide a test of the null hypothesis $H_0 : OT \not\succ_2 IT$. The test statistic is the same as in DD (2000), except that instead of the extremal T-statistic we are now interested in the *minimal* T-statistic. This statistic is computed for the values of z that are sample points within the restricted interval, i.e. in this interval we have coupled log-transformed observations of W_{IT} and W_{OT}. As opposed to the DD (2000) test, there is no restriction on the number of these points and we compute the minimal value of $\hat{T}(z)$, in the restricted interval.[23] If the minimal value is negative, the null of non-dominance is accepted. Otherwise, there exists a bootstrap approach for the derivation of the p-values for the null

[23] It may be shown that $\hat{T}(z)$ is monotonic between the sample points; therefore the minimal value of

hypothesis, which is described in detail in DD (2006) and Davidson (2007). In our tests, we use 999 bootstrap replications in order to derive the *p*-values in the tables.

There is a cost in adopting the DD (2006) null, because, as it can be analytically shown, this null cannot be rejected over the entire support of the sample distribution. DD (2006) overcame this problem by restricting the interval over which the null may be rejected to the interior of the support, excluding points at the edges. They then showed by simulation that inferences on the basis of this restricted interval constitute the most powerful available inference on the existence of stochastic dominance. We follow their suggestion on the method for restricting the interval, which we also test on simulated data.[24]

Appendix E: Conditional versus unconditional tests

For each time point of our sample we generate artificially samples of stock return paths drawn from a bootstrapped distribution constructed from the (approximately 22) observed daily stock returns till option expiration *for each one* of the 247 dates t=1,…, 247 in our data period. Such a distribution represents the information that the trader would have used to estimate the bounds had she been able to observe it. For each stock return path of the bootstrap we then compute the wealth indices for OT and IT, W_{OT} and W_{IT}, and generate two distributions of these quantities at each date in our period and for each method of estimating the bounds; recall that all these estimation methods use only quantities that can be observed by the trader before adopting her option position. Hence, evidence that the returns of the option trader dominate the returns of the index trader for "most" of the time points of our sample validates the use of the observable distribution for estimating the bounds in lieu of the unobservable distribution of the actual stock return paths.

Since these bootstrapped samples are large, we can treat the samples as the entire populations, applying a direct stochastic dominance test based on the integral condition that defines stochastic dominance.[25] This integral condition takes the following form, for $J = OT,\ IT$ and for $\underline{z}$ denoting the lower limit of the joint support of the two distributions

---

$\hat{T}(z)$ may be found only at a sample point.

[24] Details on the restrictions of the test interval are available from the authors on request.

[25] See, for instance, Hanoch and Levy (1969). Condition (9) can be easily shown, through integration by

$$D_{OT}^2(z) - D_{IT}^2(z) \geq 0, \text{ where } D_J^2(z) = \int_{\underline{z}}^{z} (z - x)\, dF_J(x) \tag{E.1}$$

In the particular case in which we observe the paired wealth levels W_{OT} and W_{IT} from a sample of size N with values W_{Ji}, $i = 1,...,N$, $J = OT, IT$ the test statistic $D_J^2(z)$ becomes

$$D_J^2(z) = \frac{1}{N}\sum_{i=1}^{N}(z - W_{Ji})_+ \tag{E.2}$$

For the bootstrapped distribution, we calculate the SD2 test statistic from (E.1)-(E.2) for the hypothesis $H_0 : OT_t \succ_2 IT_t$ t=1,…, 247 as above and decide on acceptance/rejection at a chosen significance level α (say 5%). Next, we set the variable Z_t to equal one if $\text{Prob}\{H_0 \text{ false}\} \leq \alpha$, and zero otherwise. The hypothesis $\Pr\text{ob}\{OT_t \succ IT_t\} > 0.5$ for any *t*, against the alternative $\Pr\text{ob}\{OT_t \succ_2 IT_t\} \leq 0.5$ for any *t*, is accepted if $\sum_1^{248} Z_t \geq \beta$, where β is chosen according to the desired significance level from the binomial distribution with probability p = ½ .

In Table 9, panel A, we present the results of these conditional tests. The upper panel tests the hypothesis $\Pr\text{ob}\{OT_t \succ IT_t\} > 0.5$ for the observed option bid prices that violate the call and put upper bounds in equations (1)-(4) under the same conditions as Table 2. The results are strongly supportive of the null hypothesis in all but one case for which there are too few observations, and are in full agreement with the results of the unconditional test of Table 2. Similar results also hold for the options that violate the option upper bounds under the conditions of Table 3, with the basis risk set equal to 0.[26]

In Table 9, panel B, we present the results of tests of the hypothesis $\Pr\text{ob}\{OT_t \succ IT_t\} > 0.5$ for the artificial set of options written at the upper bounds of the call and put options, as in the upper panel of Table 6. Again, the results are strongly supportive of the null hypothesis in all cases, with the observed probabilities $\Pr\text{ob}\{OT_t \succ IT_t\}$ greater

parts, to be equivalent to the better-known form of the integral condition used in most SD studies.
[26] The results are available from the authors on request.

than 65% in all but one case and always significantly greater than 50%.[27] Hence, conditional and unconditional tests agree here as well. Similar results (available upon request) establish the validity of the hypothesis $\Pr\text{ob}\{OT_t \succ IT_t\} > 0.5$ for call options purchased at the lower bound of equation (6), while for put options purchased at the lower bound of equations (4) and (5) the hypothesis is verified in all cases except for $K/F < 0.98$, again as in the unconditional tests.

[27] Similar, although slightly weaker, results also hold for the option upper bounds for the case where there is no basis risk in computing the bounds.

References

Black, F. and M. S. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, 637-654.

Bogle, J. C., 2005, "The Mutual Fund Industry 60 Years Later: For Better or Worse?" *Financial Analyst Journal* 61, 15-24.

Cerny, A., 2004, "Introduction to Fast Fourier Transform in Finance," *Journal of Derivatives* 12, 73-88.

Constantinides, G. M., 1986, "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy* 94, 842-862.

Constantinides, G. M., 2002, "Rational Asset Prices," *Journal of Finance* 57, 1567-1591.

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis, 2008, "Mispricing of S&P 500 Index Options," *Review of Financial Studies*, forthcoming.

Constantinides, G. M. and S. Perrakis, 2002, "Stochastic Dominance Bounds on Derivatives Prices in a Multiperiod Economy with Proportional Transaction Costs," *Journal of Economic Dynamics and Control* 26, 1323-1352.

Constantinides, G. M. and S. Perrakis, 2007, "Stochastic Dominance Bounds on American Option Prices in Markets with Frictions," *Review of Finance* 11, 71-115.

Coval J. D. and T. Shumway, 2001, "Expected Option Returns," *Journal of Finance* 56, 983-1009.

Davidson, R., 2007, "Testing for Restricted Stochastic Dominance: Some Further Results," Working Paper, McGill University.

Davidson, R. and J.-Y. Duclos, 2000, "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica* 68, 1435-1464.

Davidson, R. and J.-Y. Duclos, 2006, "Testing for Restricted Stochastic Dominance," Working Paper, McGill University.

Dimson, E., P. Marsh, and M. Staunton, 2006, "The Worldwide Equity Premium: A Smaller Puzzle," Working paper, London Business School.

Fama, E. F., and K. R. French, 2002, "The Equity Premium," *Journal of Finance* 57, 637-660.

Härdle, W., 1990, *Applied Nonparametric Regression*, Cambridge Univ. Press.

Hull, J. C., 2006, *Options, Futures, and Other Derivatives*, Prentice Hall.

Jackwerth, J. C., 2000, "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies* 13, 433-451.

Jackwerth, J. C., 2004, "Option-Implied Risk-Neutral Distributions and Risk Aversion," ISBN 0-943205-66-2, Research Foundation of AIMR, Charlottesville, USA.

Jackwerth, J. C. and M. Rubinstein, 1996, "Recovering Probability Distributions from Option Prices," *Journal of Finance* 51, 1611-1631.

Jobson J. D. and B. M. Korkie, 1981, "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *Journal of Finance* 4, 889-908.

McDonald, R. L., 2006, *Derivatives Markets*, Addison-Wesley.

Memmel, C., 2003, "Performance Hypothesis Testing with the Sharpe Ratio," *Finance Letters* 1, 21-23.

Merton, R. C., 1973, "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science* 4, 141-183.

Perrakis, S. and M. Czerwonko, 2006, "Transaction Costs and Stochastic Dominance Efficiency in the Index Futures Options Market," Working paper, Concordia University.

Rubinstein, M., 1994, "Implied Binomial Trees," *Journal of Finance* 49, 771-818.

Stoline, M. R. and H. K. Ury, 1979, "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons among Means," *Technometrics* 21, 87-93.

Yu, K. and M. C. Jones, 1998, "Local Linear Quantile Regression," *Journal of the American Statistical Association* 93, 228-237.

Table 1

Prediction Error of Monthly Volatility, 1983-2006

Prediction mode	Mean	Median	St. dev.	Skew.	Ex. Kurt.
Unconditional	0.0429	0.0649	0.0680	-1.7300	3.8296
90-day	0.0095	0.0076	0.0595	0.2687	5.2490
Adjusted IV	-0.0005	0.0002	0.0496	-0.2625	3.4680
GARCH	0.0177	0.0185	0.0531	0.0936	7.8302

The errors are defined as the difference between the monthly volatility and the volatility predicted by a given mode. The unconditional volatility is the sample standard deviation over the period January 1928 to January 1983. The 90-day volatility is the sample standard deviation over the preceding 90 trading days. The adjusted IV is the ATM IV on the preceding day, adjusted by the mean prediction error for all dates preceding the given date, where we drop from the preceding days all 21 pre-crash observations. The GARCH volatility is the volatility using GARCH coefficients estimated for S&P 500 daily returns over January 1928 to January 1983 and applied to residuals observed over the 90 days preceding each sample date to form projections of the volatility realized till the option expiry date.

Table 2
Returns of Options Trader and Index Trader

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) *p*-value		DD (2006) *p*-value
			$H_0: OT \succ_2 IT$	$H_0: IT \succ_2 OT$	$H_0: OT \not\succ_2 IT$
			A: Call Upper Bound		
Unconditional	43 (247)	0.090	>0.1	<0.01	0
90-day	101 (247)	0.062	>0.1	<0.01	0
Adjusted IV	120 (226)	0.086*	>0.1	<0.01	0
GARCH	65 (247)	0.093	>0.1	<0.01	0
			B: Put Upper Bound		
Unconditional	23 (247)	0.128*	>0.1	>0.1	0
90-day	16 (247)	0.007	>0.1	>0.1	0.078
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a
GARCH	9 (247)	n/a	n/a	n/a	n/a
			C: Put Upper Bound Implied by Call Upper Bound		
Unconditional	39 (247)	0.074	>0.1	<0.05	0
90-day	68 (247)	0.062	>0.1	<0.01	0
Adjusted IV	74 (226)	0.152**	>0.1	<0.01	0
GARCH	49 (247)	0.141**	>0.1	<0.01	0

Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols * and ** denote a difference in the Sharpe ratios significant at the 10% and 5% levels, respectively, in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0: IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with $k = 20$ and $\nu = \infty$.

Table 3

A: Returns of Straddles Trader and Index Trader

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$			DD (2000) *p*-value		DD (2006) *p*-value
		Straddle	Call	Put	$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \nsucc_2 IT$
Unconditional	34 (247)	0.270***	0.179**	0.204**	>0.1	>0.1	0.032
90-day	66 (247)	0.178**	0.077	0.112*	>0.1	>0.1	0.001
Adjusted IV	73 (226)	0.357***	0.202**	0.227***	>0.1	<0.05	0.003
GARCH	41 (247)	0.365***	0.177**	0.211***	>0.1	>0.1	0.028

B: Returns of Strangles Trader and Index Trader

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$			DD (2000) *p*-value		DD (2006) *p*-value
		Straddle	Call	Put	$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \nsucc_2 IT$
Unconditional	40 (247)	0.346***	0.153**	0.265***	>0.1	>0.1	0.010
90-day	81 (247)	0.270***	0.074	0.206***	>0.1	<0.05	0.000
Adjusted IV	92 (226)	0.400***	0.165**	0.324***	>0.1	<0.05	0.003
GARCH	56 (247)	0.389***	0.141**	0.271***	>0.1	>0.1	0.007

Equally weighted average of all violating options equivalent to one call and one put per share was traded at each date. Trades were executed whenever there was a call violating the upper bound and a put traded at the same strike for the same date. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% level, respectively. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and $\nu = \infty$.

Table 4
Returns of Options Trader and Index Trader—without Futures Basis Risk

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value		DD (2006) p-value
			$H_0: OT \succ_2 IT$	$H_0: IT \succ_2 OT$	$H_0: OT \not\succ_2 IT$
A: Call Upper Bound					
Unconditional	68 (247)	0.045	>0.1	<0.01	0
90-day	157 (247)	0.078^{*}	>0.1	<0.01	0
Adjusted IV	195 (226)	0.084^{**}	>0.1	<0.01	0
GARCH	112 (247)	0.100^{*}	>0.1	<0.01	0
B: Put Upper Bound					
Unconditional	36 (247)	0.050	>0.1	<0.05	0
90-day	52 (247)	0.051	>0.1	<0.01	0
Adjusted IV	64 (226)	0.074	>0.1	<0.01	0
GARCH	38 (247)	0.101^{*}	>0.1	<0.01	0

The table differs from Table 2 only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the risk by $\bar{\varepsilon} = 0.5$. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols * and ** denote a difference in the Sharpe ratios significant at the 10% and 5% level, respectively, in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0: IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and $\nu = \infty$.

Table 5
Returns of Straddles Trader and Index Trader—without Futures Basis Risk

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$			DD (2000) p-value		DD (2006) p-value
		Straddle	Call	Put	$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \not\succ_2 IT$
Unconditional	50 (247)	0.187^{**}	0.117^{*}	0.143^{**}	>0.1	>0.1	0.020
90-day	132 (247)	0.178^{**}	0.093^{*}	0.116^{**}	>0.1	<0.01	0
Adjusted IV	166 (226)	0.230^{**}	0.184^{***}	0.208^{***}	>0.1	<0.01	0
GARCH	99 (247)	0.315^{***}	0.190^{***}	0.235^{***}	>0.1	<0.1	0

The table differs from Table 3 only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the risk by $\bar{\varepsilon} = 0.5$. Equally weighted average of all violating options equivalent to one call and one put per share was traded at each date. Trades were executed whenever there was a call violating the upper bound and a put traded at the same strike for the same date. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% level, respectively. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and $\nu = \infty$.

Table 6
Returns of Options Trader and Index Trader—with Risk Aversion Coefficient 10

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value		DD (2006) p-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \nsucc_2 IT$
			A: Call Upper Bound		
Unconditional	43 (247)	0.184*	>0.1	<0.01	0
90-day	101 (247)	0.161	>0.1	<0.01	0
Adjusted IV	120 (226)	0.222	>0.1	<0.01	0
GARCH	65 (247)	0.178	>0.1	<0.01	0
			B: Put Upper Bound		
Unconditional	23 (247)	0.231	>0.1	>0.1	0.010
90-day	16 (247)	0.062	>0.1	>0.1	0.094
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a
GARCH	9 (247)	n/a	n/a	n/a	n/a

The table differs from Table 2 only in that the risk aversion coefficient is set to 10, instead of 2. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with $k = 20$ and $\nu = \infty$.

Table 7

Returns of Options Trader and Index Trader—with Equity Risk Premium 6%

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value		DD (2006) p-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \not\succ_2 IT$
			A: Call Upper Bound		
Unconditional	38 (247)	0.068	>0.1	<0.01	0
90-day	86 (247)	0.045	>0.1	<0.01	0
Adjusted IV	96 (226)	0.068	>0.1	<0.01	0
GARCH	58 (247)	0.083	>0.1	<0.01	0
			B: Put Upper Bound		
Unconditional	23 (247)	0.117*	>0.1	>0.1	0
90-day	11 (247)	0.008	>0.1	>0.1	0.252
Adjusted IV	3 (226)	n/a	n/a	n/a	n/a
GARCH	6 (247)	n/a	n/a	n/a	n/a

The table differs from Table 2 only in that the risk premium is set to 6%, instead of 4%. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and $\nu = \infty$.

Table 8
Returns of Options Trader and Index Trader—without the Crash Period

Volatility prediction mode	# months with viol. (#months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value		DD (2006) p-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \not\succ_2 IT$
		A: Call Upper Bound			
Unconditional	38 (241)	0.066	>0.1	<0.01	0
90-day	100 (241)	0.074	>0.1	<0.01	0
Adjusted IV	119 (220)	0.081	>0.1	<0.01	0
GARCH	61 (241)	0.082	>0.1	<0.01	0
		B: Put Upper Bound			
Unconditional	19 (241)	0.085	>0.1	>0.1	0.081
90-day	16 (241)	0.007	>0.1	>0.1	0.078
Adjusted IV	4 (220)	n/a	n/a	n/a	n/a
GARCH	9 (241)	n/a	n/a	n/a	n/a

This table differs from Table 2 only in that the seven observations which include the date of the October crash and the following six months were excluded. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbol * denotes a difference in the Sharpe ratios significant at the 10% level in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \not\succ_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and $\nu = \infty$.

Table 9
Returns of Options Trader and Index Trader—Non-Stationary Distribution

Panel A: Observed Option Prices

Volatility prediction mode	Call Upper Bound		Put Upper Bound	
	# months with viol. (# months)	Proportion $OT \succ_2 IT$	# months with viol. (# months)	Proportion $OT \succ_2 IT$
Unconditional	43 (247)	0.609**	23 (247)	0.890***
90-day	101 (247)	0.634***	16 (247)	0.806***
Adjusted IV	120 (226)	0.649***	4 (226)	n/a
GARCH	65 (247)	0.608**	9 (247)	n/a

Panel B: Option Prices on the Bounds

Volatility prediction mode	Call Upper Bound	Put Upper Bound
	Proportion $OT \succ_2 IT$	Proportion $OT \succ_2 IT$
Unconditional	0.869***	0.949***
90-day	0.813***	0.969***
Adjusted IV	0.799***	0.958***
GARCH	0.856***	0.978***

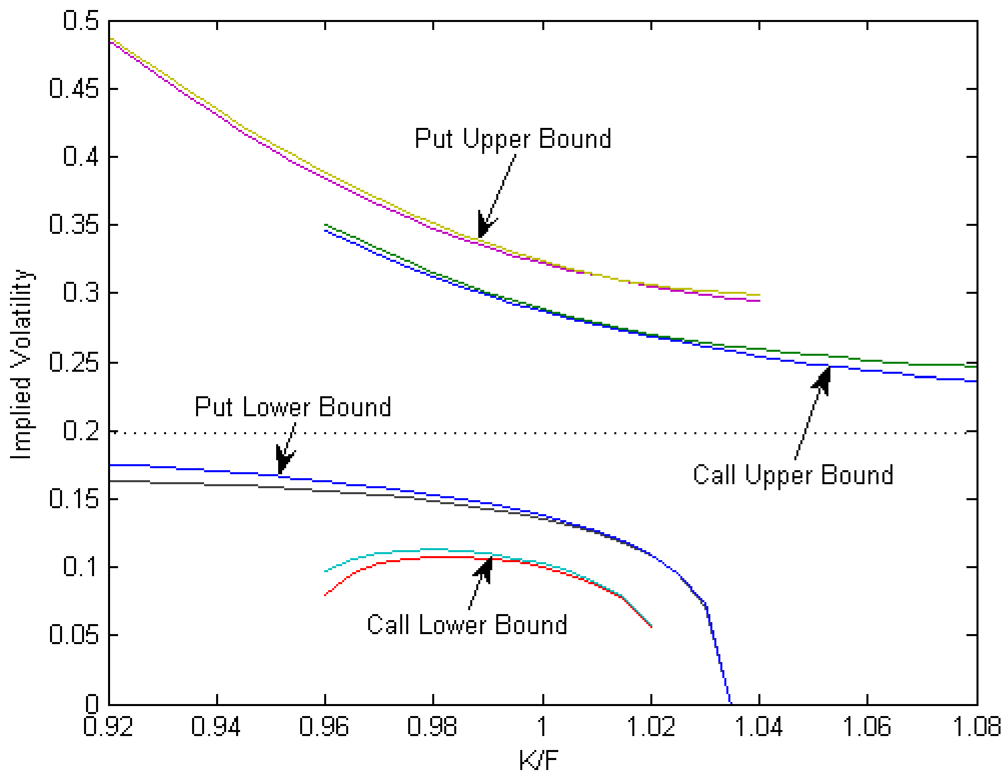
This table shows the proportion of stochastic dominance tests in which the conditional bootstrapped distribution of the option trader's wealth dominated that of the index trader, as described in Appendix E. In panel A, the set of mispriced options is the same as in Table 2. In Panel B, the options are written with a price equal to the corresponding bound. The significance levels are for a binomial sign test that the indicated proportion exceeds 50%. The symbols ** and *** indicate significance respectively at 5% or better and 1% or better.

Table 10
Returns of Naïve Trader and Index Trader

Trade type	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	DD (2000) p-value		DD (2006) p-value
			$H_0 : OT \succ_2 IT$	$H_0 : IT \succ_2 OT$	$H_0 : OT \nsucc_2 IT$
A: 90th or 10th Critical Quantile					
Short call	58 (243)	0.097	>0.1	<0.01	0
Short put	67 (243)	0.186^{**}	>0.1	<0.01	0
Long call	73 (243)	-0.229^{***}	>0.1	>0.1	0.153
Long put	95 (243)	-0.039	>0.1	>0.1	0.150
B: 97th.5 or 2nd.5 Critical Quantile					
Short call	32 (243)	0.156^{**}	>0.1	<0.01	0
Short put	36 (243)	0.215^{***}	>0.1	<0.05	0
Long call	27 (243)	-0.080	>0.1	>0.1	0.222
Long put	45 (243)	0.059	>0.1	>0.1	0.101

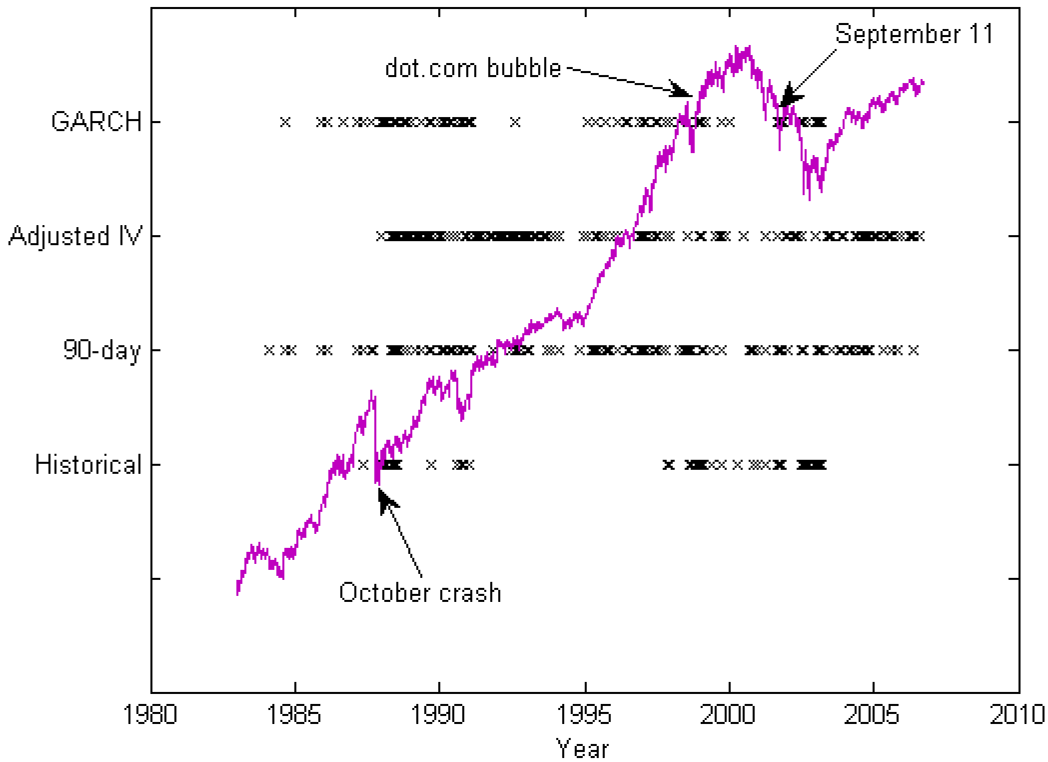
Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols ** and *** denote a difference in the Sharpe ratios significant at the 5% and 1% level, respectively, in a one-sided test. P-values for the Davidson-Duclos (2006) test are based on 999 bootstrap trials. The p-values of $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with k = 20 and $\nu = \infty$.

Figure 1: Illustration of Upper and Lower Bounds on Call and Put Options



Bound were derived for σ = 0.20 imposed on a 90-day distribution for a date in our sample. 95% CI were derived by bootstrapping the 90-day distribution and exemplify the bounds dependence on the third and fourth distribution moments.

Figure 1: Time Distribution of Observed Violations



The figure displays the violations of the call upper and put lower bounds against 247 dates with app. monthly periodicity for the period February 1983-July 2006. For the adjusted IV distributions, the first 21 dates are not in the sample. The line across the plot is the natural logarithm of the S&P 500 index.